

# Studi & Ricerche

## STUDIES & RESEARCH

Michele Lanzetta, Federica Fanti, Giovanni Tantussi\*

Le fasi principali del riconoscimento dei caratteri alfanumerici

*The main stages in alphanumerical character recognition*

Il preprocessing delle immagini è fondamentale per l'OCR

*Preprocessing images is fundamental in OCR*

Metodi di sogliatura e tecniche morfologiche principali

*Main thresholding and morphing techniques*

# Automazione e controllo visivo della marcatura di prodotti (Parte Prima)

## *Automating and visually inspecting product marking (Part One)*

**L**e crescenti esigenze di identificazione e rintracciabilità (ISO 9000) trovano spesso una soluzione nella marcatura di codici sui prodotti o sugli imballaggi.

Tra i metodi più noti ricordiamo i *data matrix*, i codici a barre e i codici alfanumerici. Mentre i primi due tipi, che vengono solitamente stampati su confezionamenti e imballaggi, prevedono metodi standardizzati per la lettura, la variabilità dei codici alfanumerici pone problemi dovuti ad esempio ai diversi tipi di carattere (*font*), senza parlare del gran numero di cifre, lettere, simboli, accenti e segni di interpunzione presenti nelle varie lingue.

Le diverse tecnologie di marcatura pongono poi ulteriori difficoltà dovute ad esempio al contrasto del tratto rispetto allo sfondo (si pensi ad esempio allo *stencil* fatto con vernici su container) o ai problemi dovuti alla irregolarità del tratto (si consideri il caso di caratteri micropunzonati o la stampa ad aghi).

In questo articolo, suddiviso in due parti (di cui la seconda

**G**rowing demands for identification and traceability (ISO 9000) often find a solution in marking codes on products or packages. Among the best-known methods we cite data matrix, barcodes and alphanumerical codes. While the first two types, usually printed on packages and packing, use

standard scanning methods, the variability of alphanumerical codes poses problems due, for example, to different types of lettering (fonts), not to mention the great number of ciphers, letters, symbols, accents and punctuation signs present in various languages. And different marking technologies create further difficulties due, for instance, to contrast between the marking and the background (one example is stenciling done with paint on containers) or to problems caused by uneven marking (such as micro-punched letters or dot matrix printing).

This article, divided into two parts (the second of which will be published in the next issue of the magazine), is

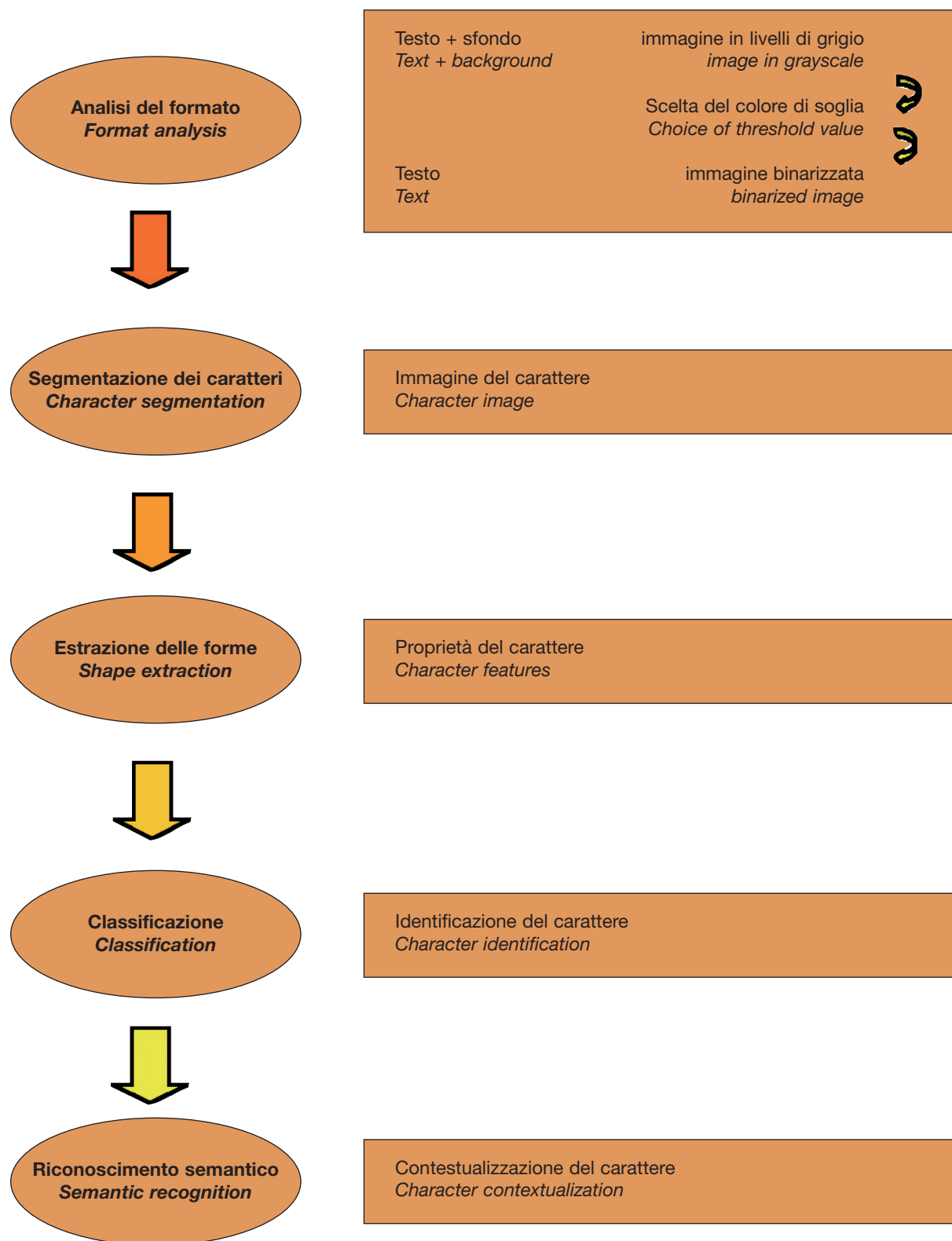


Fig. 1. Sequenza delle principali fasi presenti nei sistemi di riconoscimento caratteri.  
*Sequence of the main stages in character recognition systems.*

# Studi & Ricerche

## STUDIES & RESEARCH

sarà pubblicata sul prossimo numero della rivista), verranno indicate le principali fasi richieste per il controllo visivo della marcatura.

Tali fasi sono state estrapolate dall'ampia letteratura scientifica sull'argomento per cui sono stati anche forniti i termini tecnici anglosassoni, utilizzabili per eventuali approfondimenti.

Non viene riportata una letteratura specifica tenuto conto della vastità della tematica, che si può esemplificare con poche cifre: oltre 1500 *record* (pubblicazioni) presenti nel database bibliografico Compendex [1] e oltre 2800 su 1.574.558 del database IEEE [2]. Una finalità di questo articolo è infatti quella di fornire alcune iniziali indicazioni per orientarsi su questa affermata tecnologia.

Le varie operazioni necessarie per il riconoscimento dei caratteri vengono **implicitamente** applicate da un eventuale addetto al controllo del prodotto nel caso di sistemi di produzione manuali, mentre devono essere implementate, tramite appositi algoritmi, per lo sviluppo di sistemi automatici.

Il riconoscimento dei caratteri con metodi ottici è la fase più nota dell'intero processo che ne comprende altre precedenti e successive, come mostrato in figura 1.

### L'Optical Character Recognition (OCR)

Le principali fasi del riconoscimento di caratteri sono riportate in figura 1 e sono discusse nei paragrafi successivi.

### Pre-elaborazione (preprocessing)

A titolo di esempio si riportano alcuni esempi di elaborazioni di base.

Spesso, durante l'analisi dell'immagine, viene rilevato del rumore ad esempio *salt and pepper* o gaussiano [3] che falserebbe il risultato finale se non preso in considerazione. In tali condizioni è necessario ricorrere al filtraggio, che deve essere applicato tenendo conto del tipo di rumore che si vuole eliminare.

Successivamente il risultato può essere ulteriormente migliorato ricorrendo alle **operazioni morfologiche**, alcune delle quali sono elencate di seguito e che possono essere eseguite su immagini binarie o a colori:

- *Dilatazione* per estendere una regione dell'immagine;
- *Erosione* per ridurre una regione dell'immagine;
- *Closing* e *opening* cioè una dilatazione seguita da una erosione e viceversa per unire due regioni separate o per eliminare piccoli elementi isolati;
- *Thinning* per individuare il profilo medio del carattere.

Alcune elaborazioni sono illustrate nelle figure 2 e 3.

Esistono tecniche più sofisticate per operare sulle immagini a livelli di grigio (*grayscale*) [4].

Nel caso della marcatura di prodotti, i disturbi dovuti alla disuniformità del materiale o all'imperfezione della tecnologia stessa possono richiedere elaborazioni aggiuntive delle immagini acquisite. Si osservi ad esempio la figura 4: i singoli caratteri sono confusi; si riconosce il testo grazie alle proprie conoscenze pregresse (*context learning*).



Fig. 2.  
a. Immagine originale di una copertina di un libro a colori.

a. Original image of a book cover in color.



b. Operazione di *closing* ottenuta applicando alle zone chiare erosioni di 8 pixel dopo altrettante dilatazioni su un'immagine binarizzata (bianco e nero). I caratteri più piccoli del sottotitolo si sono fusi in una macchia (*blob*) indistinta più grande. Il testo "seconda edizione" scritto con carattere scuro su sfondo chiaro è sparito. I caratteri più grandi cominciano a fondersi tra loro.

b. Closing, by applying 8-pixel erosions to the pale zones after an equal number of dilations on a binarized (black and white) image. The smaller characters in the subtitle have merged into a blob. The words "seconda edizione" (second edition) written in dark letters on a light background have disappeared. The larger letters begin to run together.



c. Operazione di *opening* ottenuta applicando alle zone chiare dilatazioni di 10 pixel dopo altrettante erosioni. La figura e i caratteri minori sono quasi spariti. Il testo maggiore è rimasto inalterato. Anche dai contorni della figura sono stati eliminati i piccoli elementi di disturbo chiari.

c. Opening, achieved by applying 10-pixel dilations after an equal number of erosions. The illustration and smaller letters have nearly disappeared. The larger letters are unchanged. Even the little pale noise elements have been eliminated from the outlines of the illustration.

# Studi & Ricerche

## STUDIES & RESEARCH

### TECNOLOGIA MECCANICA

e studi  
di fabbricazione

SECONDA EDIZIONE

Fig. 3.

a. Esempio di immagine binaria 460x230 pixel con rumore (riflessi).

a. Example of a binary 460x230 pixel image with noise (reflections).

### TECNOLOGIA MECCANICA

e studi  
di fabbricazione

SECONDA EDIZIONE

b. Esempio di immagine dopo erosione con elemento strutturale quadrato di lato 7 pixel. La maggior parte del rumore è stata rimossa.

b. Example of an image after erosion with a 7-pixel square structural element. Most of the noise has been removed.

### TECNOLOGIA MECCANICA

e studi  
di fabbricazione

c. Esempio di immagine dopo erosione con elemento strutturale quadrato di lato 15 pixel. Il rumore è stato quasi completamente eliminato intorno ai caratteri più grandi del titolo, che possono facilmente essere "segmentati" (separati) rispetto allo sfondo. Sono praticamente spariti i caratteri di dimensione inferiore.

c. Example of the image after erosion with a 15 - pixel square structural element. The noise has been almost completely eliminated from around the largest letters in the title, which can be easily "segmented" (separated) from the background. The smaller letters have almost completely disappeared.

concerned with the main stages required for visually inspecting marking.

These stages have been extrapolated from the broad range of scientific literature on the subject but we do not cite specific texts, given the vastness of the theme: for example, there are over 1500 records in the Compendex bibliographical database [1] and more than 2800 out of 1,574,558 in the IEEE database [2]. One purpose of this article is, in fact, to provide initial notions for orientation in this well-known technology.

The various operations involved in character recognition are **implicitly** performed by an employee assigned to product inspection in the case of manual systems, while they require special algorithms for the development of automatic systems.

Character recognition with optical methods is the best-known stage in the entire process, which also includes prior and subsequent ones as shown in figure 1.

### Optical Character Recognition (OCR)

The main stages of character recognition are given in figure 1 and are discussed in subsequent paragraphs.

### Preprocessing

We give some basic examples.

Often, during image analysis, salt and pepper or Gaussian [3] noises are detected that would falsify the final result unless they are taken into consideration. Under these conditions there is a need for filtering, which has to be done taking into account the type of noise to be eliminated. Afterwards the result can be further improved through **morphological operations** (morphing), some of which are listed here and can be done on binary or color images:

- Dilation to expand a portion of the image;
- Erosion to reduce a portion of the image;
- Closing and opening, that is, dilation and then erosion and vice-versa to unite two separate portions or eliminate little isolated elements;
- Thinning to individuate the average profile of the character.

Some processing types are shown in figures 2 and 3.

There are more sophisticated techniques to work on images in grayscale [4].

In the case of product marking, noise coming from uneven material or imperfect technology may require additional processing of the acquired images. Observe figure 4, for example: the individual characters are confused; the text is recognized through former knowledge (context learning).

**Thresholding** techniques [5] are suitable for scanning printed characters on paper supports that represent most of the characters in existence. Applications regard postal services, printed matter (books, newspapers, periodicals) and include the special case of handwritten characters.

The search for threshold value, that is, the minimum grayscale value of the histogram, plays a fundamental role [6] and in fact it is right in this part of the system that error



# Studi & Ricerche

## STUDIES & RESEARCH

Le tecniche di **sogliatura** (*thresholding*) [5] si prestano per la lettura di caratteri stampati su supporti cartacei che rappresentano la maggior parte dei caratteri esistenti. Applicazioni riguardano i servizi postali, la carta stampata (libri, giornali, riviste) e comprendono il caso particolare dei caratteri manoscritti (*handwritten*).

La ricerca del valore della soglia, cioè il valore minimo nell'istogramma dei livelli di grigio, ha un ruolo fondamentale [6], infatti è proprio in questa parte del sistema che la propagazione degli errori è consistente, quindi è d'obbligo costruire un metodo di scelta della soglia robusto.

Se i pixel dello sfondo e del testo in primo piano hanno valori abbastanza costanti rispetto all'intera immagine, allora è possibile trovare un singolo valore di soglia per l'immagine. Per decidere come operare sull'immagine la letteratura scientifica propone due approcci [7], chiamati rispettivamente metodo di **soglia globale** (*global threshold*) e metodo di **soglia locale** (*local threshold*). Tra i due il più semplice e immediato è il metodo di soglia globale in cui viene usata una singola soglia per tutti i pixel dell'immagine. Non è detto però che per diverse immagini sia corretto applicare un'unica soglia, come accade se i pixel dello sfondo e quelli del testo in primo piano non hanno valori uniformi.

### Global thresholding

La strada più diretta per selezionare una soglia nel metodo globale è l'uso dell'istogramma dei livelli di grigio. Nel diagramma (figura 5) ad ogni valore delle intensità dei pixel dell'immagine viene associato il numero di pixel che corrispondono a quel valore di intensità. Per un'immagine che ha l'intensità dei pixel dello sfondo ben diversa da quella dei pixel della parte in primo piano, il diagramma risulta avere due picchi ben distinti (bimodale). La valle che si trova fra questi due è considerata il minimo tra i due massimi e il valore dell'intensità viene scelto come valore di soglia globale che meglio separa i due picchi. Un metodo popolare, conosciuto fin dagli anni '80, è il metodo di Otsu [8]. Prima di tutto viene eseguito il calcolo del rapporto fra la variazione fra le classi e la variazione all'interno delle classi stesse per ogni potenziale valore di soglia. Ogni

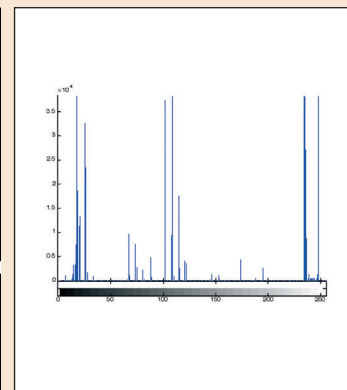


Fig. 5. Binarizzazione dell'immagine in figura 2 con il metodo di Otsu (sinistra) e istogramma dei livelli di grigio (destra).

*Binarizing the image in figure 2 with the Otsu method (left) and the grayscale histogram (right).*

propagation is heavy, making it necessary to create a robust method for threshold choosing.

If the pixels of the background and of the foreground text have values fairly constant with respect to the entire image then a single threshold value can be found for the image. To decide how to work on the image, scientific literature offers two approaches [7] called, respectively, the **global threshold** and **local threshold** methods. Of the two the simplest and most immediate is the global threshold method, using a single threshold for all the image's pixels. But if the background pixels and those of the foreground text do not have uniform values, applying a single threshold may not necessarily be correct.

### Global thresholding

The straightest path to threshold selection in the global method is using the grayscale histogram. In the diagram (figure 5) every intensity value of the image's pixels is associated with the number of pixels corresponding to that same level of intensity. For an image in which the background's pixel intensity is very different from pixel intensity in the foreground, the diagram will have two very distinct (bimodal) peaks. The valley between them is considered the minimum between two maximums and the



Fig. 4. Immagine acquisita da fotocamera (sinistra). Conversione in livelli di grigio (centro). Immagine binarizzata in cui compare il testo (destra).  
*Image acquired by a camera (left). Conversion into grayscale (center). Binarized image in which the text appears (right).*

# Studi & Ricerche

## STUDIES & RESEARCH

classe è costituita dai pixel dello sfondo e della parte in primo piano che hanno lo stesso valore di intensità di livello di grigio. L'obiettivo è trovare la soglia che massimizza la variazione di intensità fra le (due) classi e che la minimizza all'interno delle classi stesse, in modo tale da avere il massimo contrasto dell'immagine e quindi la massima leggibilità.

Diversi approcci per determinare la soglia di binarizzazione sono descritti in [9] e [10].

### Local thresholding

Un modo comune di utilizzare il metodo locale è l'analisi delle intensità dei livelli di grigio nella zona locale dell'immagine considerata, per determinare i valori delle soglie locali. In seguito diversi algoritmi possono essere usati per cambiare dinamicamente il valore di soglia. Uno dei primi problemi del metodo locale è la scelta della grandezza della finestra da considerare per la determinazione del valore di soglia. Infatti la finestra scelta deve essere abbastanza grande da garantire che il numero di pixel dello sfondo sia sufficiente per ottenere una buona stima del valore medio, ma non così grande che la media risulti troppo diversa dall'intensità locale. Uno degli algoritmi più popolari è quello di Niblack [11]. Questo valore di soglia è basato sul calcolo della media locale e della deviazione standard. La soglia nasce dalla formula seguente:

$$T(x, y) = m(x, y) + k * s(x, y)$$

dove  $m(x, y)$  e  $s(x, y)$  sono rispettivamente l'intensità media nell'area locale e il valore della deviazione standard.  $k$  è una costante correttiva che è usata per aggiustare il valore del risultato in dipendenza del tipo di immagine. Questo è un altro svantaggio del metodo locale, infatti alcuni di questi metodi hanno molti parametri da regolare che non possono essere estesi ad un'ampia gamma di immagini. Inoltre gli algoritmi possono deteriorare in diversi modi i caratteri in una stessa parola e quindi spesso non sono convenienti per la fase di riconoscimento.

Zhang e Tan [12] propongono una versione rivista della formula di Niblack:

$$T(x, y) = m(x, y) * [1 + k * (1 - s(x, y) / R)]$$

dove  $k$  e  $R$  sono ancora costanti empiriche, ma permettono di ridurre la sensibilità del sistema al rumore.

Le prestazioni degli algoritmi del metodo globale sono state valutate in alcuni studi usando misure di uniformità e forma del carattere ottenuto dopo sogliatura [13]. Fra oltre 20 algoritmi di calcolo delle soglie, è stato concluso che i metodi più soddisfacenti sono: il metodo di separazione delle classi di Otsu [8], il metodo di conservazione del momento di Tsai, il metodo dell'entropia di Johannsen e Bille e quello dell'entropia di Kapur et al. [10]; l'ordine di successo dipende comunque dall'immagine analizzata.

*amount of intensity is chosen as the global thresholding value separating the two peaks. A popular method known from the 1980s is the Otsu method [8]. First of all a calculation is made of the ratio between variation between classes and variation within the classes themselves for each potential threshold value. Each class is made up of the background pixels and the foreground portion having the same grayscale intensity value. The aim is to find the threshold that maximizes the variation in intensity between the (two) classes and that minimizes it within the classes, in order to get maximum image contrast and therefore maximum legibility. Different approaches to determining the binarization threshold are described in [9] and [10].*

### Local thresholding

*A common method of using the local method is analyzing grayscale intensities in the local zone of the given image to determine the values of local thresholds. Afterwards various algorithms can be used to dynamically change the threshold value. One of the first problems with the local method is choosing the size of the window to consider for determining threshold value. In fact, the chosen window must be big enough to guarantee that the number of background pixels is enough to get a good estimate of average value, but small enough for the average not to differ too much from local intensity. One of the most popular algorithms is Niblack's [11]. This threshold value is based on calculating the local average and standard deviation. The threshold comes from the following formula:*

$$T(x, y) = m(x, y) + k * s(x, y)$$

*in which  $m(x, y)$  and  $s(x, y)$  are, respectively, the mean intensity in the local area and the value of standard deviation;  $k$  is a corrective constant used to adjust result value depending on the type of image. This is another disadvantage of the local method, and in fact some of these methods have many parameters to be regulated that cannot be extended to a wide range of images. In addition, the algorithms can in many ways deteriorate the characters in the same word and so often are inconvenient in the recognition stage.*

*Zhang and Tan [12] propose a revised version of Niblack's formula:*

$$T(x, y) = m(x, y) * [1 + k * (1 - s(x, y) / R)]$$

*in which  $k$  and  $R$  are still empirical constants but make it possible to reduce the system's sensitivity to noise.*

*Algorithm performance in the global method has been assessed in a number of studies using uniformity measurements and character shape gotten after thresholding [13]. Among more than 20 algorithms for threshold calculation it was concluded that the most satisfactory methods are: Otsu's classifier [8], Tsai's moment conservation method, Johannsen and Bille's entropy method and that of Kapur et al. [10], with the amount of success in any case depending on the image analyzed.*

# Studi & Ricerche

---

## STUDIES & RESEARCH

### Note

\* Michele Lanzetta<sup>1,2</sup>, Federica Fanti<sup>1</sup>, Giovanni Tantussi<sup>1</sup>:

1. Università di Pisa, Dipartimento di Ingegneria Meccanica, Nucleare e della Produzione (DIMNP), Sezione Produzione, via Bonanno Pisano 25/B, 56126 Pisa, Italy, [lanzetta@unipi.it](mailto:lanzetta@unipi.it), tel. 050-9130.22;
2. Consiglio Nazionale delle Ricerche (CNR), Istituto di Scienze e Tecnologie dell'Informazione (ISTI) "A. Faedo", via G. Moruzzi 1, 56124 Pisa, Italia.

### Ringraziamenti

Il presente lavoro è stato sviluppato come progetto didattico di Federica Fanti nell'ambito del corso di Automazione dei Processi Produttivi del corso di laurea specialistica in Ingegneria dell'Automazione, anno accademico 2006-07, presso la facoltà di Ingegneria dell'Università di Pisa.

Michele Lanzetta e Giovanni Tantussi sono docenti del settore scientifico disciplinare Tecnologie e Sistemi di Lavorazione presso la medesima Università.

Il progetto è stato co-finanziato dal Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) e dall'Università di Pisa come Progetto di Ricerca Scientifica di Rilevante Interesse Nazionale (PRIN, 2004).

### Notes

\* Michele Lanzetta<sup>1,2</sup>, Federica Fanti<sup>1</sup>, Giovanni Tantussi<sup>1</sup>:

1. University of Pisa, Department of Mechanical, Nuclear and Production Engineering (DIMNP), Production Section, Via Bonanno Pisano 25/B, 56126 Pisa, Italy, [lanzetta@unipi.it](mailto:lanzetta@unipi.it), tel. 050-9130.22;
2. National Research Council (CNR), A. Faedo Information Technologies and Sciences Institute (ISTI), Via G. Moruzzi 1, 56124 Pisa, Italy.

### Acknowledgements

This study was developed as an educational project by Federica Fanti in the context of a course on automating production processes for the degree course in Automation Engineering, academic year 2006-07, Engineering Faculty, University of Pisa.

Michele Lanzetta and Giovanni Tantussi are professors in the scientific sector of Production Technologies and Systems at the same university.

The project was co-financed by the Ministry of Education, Universities and Research (MIUR) and by the University of Pisa as a Scientific Research Project of National Importance (PRIN, 2004).

---

### Bibliografia / Bibliography

- [1] Engineering Village, <http://www.engineeringvillage.org>, accesso di Maggio 2007 con "OCR" in subject/title/abstract.
- [2] IEEEXplore, <http://ieeexplore.ieee.org>, accesso di Maggio 2007 con "OCR" in Full Text.
- [3] M. Lanzetta, *The Introduction of a New Functionality in Existing Industrial Products: the Case of OCR in an Artificial Vision System*, Proceedings of the 10<sup>th</sup> International ADM Conference "Design Tools and Methods in Industrial Engineering", Florence (Italy), September 17<sup>th</sup>-19<sup>th</sup>, 1997, pp. 301-310, Publ. Università di Firenze, 1997, ISBN: 88 7957 113-3.
- [4] Y. Lecun, L. Bottou, Y. Bengio et al., *Gradient-based learning applied to document recognition*, Proceedings of the IEEE 86 (11): 2278-2324 Nov. 1998.
- [5] O.D. Trier, T. Taxt, *Evaluation of binarization methods for document images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994.
- [6] P.K. Sahoo, S. Soltani, A.K.C. Wong, *A Survey of Thresholding Techniques*, Computer Vision, Graphics, and Image Processing, vol. 41, pp. 233-260, 1988.
- [7] S. Mori, C.Y. Suen, K. Yamamoto, *Historical review of OCR research-and development*, Proceedings of the IEEE 80 (7): 1029-1058 Jul. 1992.
- [8] N. Otsu, *A threshold selection method from grey level histogram*, IEEE Trans. Syst. Man Cybern., vol. 9 n. 1, 1979, pp. 62-66.
- [9] Y. Liu and S. N. Srihari, *Document image binarization based on texture features*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, n. 5, pp. 540-544, May 1997.
- [10] J.N. Kapur, P.K. Sahoo, A.K.C. Wong, *A new method for graylevel picture thresholding using the entropy of the histogram*, Computer Vision, Graphics, and Image Processing, vol. 29, pp. 273-285, 1985.
- [11] W. Niblack, *An Introduction to Digital Image Processing*, pp. 115-116, Prentice Hall, 1986.
- [12] G. Leedham, C. Yan, K. Takru, J. Tan, L. Mian, *Comparison of some thresholding algorithms for text/background segmentation in difficult document images*, Proceedings of ICDAR, 2003.
- [13] O.D. Trier, A.K. Jain, *Goal-Directed Evaluation of binarization methods*, IEEE Transactions on Pattern Analysis and Machine Intelligence 17 (12): 1191-1201 Dec., 1995.